A Total Break of the 3WISE Digital Signature Scheme

Daniel Smith-Tone[1,2]

[1] University of Louisville, Louisville KY, USA
[2] National Institute of Standards and Technology, Gaithersburg, Maryland, USA
daniel.smith@nist.gov

Abstract. A new batch of "complete and proper" digital signature scheme submissions has recently been published by NIST as part of its process for establishing post-quantum cryptographic standards. This note communicates an attack on the 3WISE digital signature scheme that the submitters did not wish to withdraw after NIST communicated it to them.
While the 3WISE digital signature scheme is based on a collection of cubic maps which are naturally modeled as symmetric 3-tensors and 3-tensor rank is a difficult problem, the multivariate signature scheme is still vulnerable to MinRank attacks upon projection. We are able to break the NIST security level I parameters within a few seconds. Since the attack is polynomial time, there is no reparametrization resulting in a secure scheme.

Keywords: Multivariate Cryptography · MinRank · Cryptanalysis.

1 Introduction

About a year ago, the National Institute of Standards and Technology (NIST) published its first selections for post-quantum cryptographic standards, see [1]. Shortly thereafter, NIST published a new call for proposals for post-quantum digital signatures, see [9], citing the contrast between the relatively limited selection of secure digital signature schemes in the standardization process and the quite diverse uses of digital signatures in the world today. Now, NIST has published the "complete and proper" submissions added to the standardization process from this call.

A submission is judged to be complete and proper by passing a checklist of requirements specified in the call for proposals [9]. Usually the schemes that are deemed complete and proper are plausibly secure; however, schemes for which NIST has communicated to the submitters an attack but the submitters do not wish to withdraw from the process are still published for public analysis. The publication of such schemes does not illustrate any endorsement by NIST or any assessment of the quality of the submission; it merely indicates the satisfaction of the requirements for submission into the process. This short article presents

the attack NIST communicated with the submitter of 3WISE revealling a total and practical break of the scheme.

The 3WISE digital signature scheme [6] is a multivariate scheme utilizing cubic maps instead of the traditional quadratic maps. One of the motivations for the design of 3WISE is the fact that the natural way of representing cubic forms is by way of 3-tensors, and 3-tensor rank is a difficult problem. In fact, the central map of 3WISE is a fixed collection of 3-tensors, each of which is a rank 1 tensor.

In this work, we show that we may still apply standard MinRank techniques to attack 3WISE due to its low 3-tensor rank property. Specifically, we show that we may apply a projection to the scheme and recover a related quadratic system with a rank defect. We then show that solving the MinRank instances arising from such projections reveals information about the secret bases and powers a key recovery attack.

We implement our attack and find that we are able to efficiently recover an equivalent secret key for full scale parameter sets in seconds. Since the attack is polynomial time, there is no possible parametrization that offers a reasonable combination of security and performance.

The paper is organized as follows. First, we present the design of the 3WISE digital signature scheme. We next discuss the MinRank problem and a simple method for solving the problem that is efficient for very small target ranks. In the subsequent section, we present our attack on 3WISE and analyze its complexity. Our experimental data on our implementation of the attack on the actual parameters of 3WISE are then provided in the next section. Finally, we conclude, noting how this cryptanalysis compares with previous cryptanalyses in the literature.

2 The 3WISE Digital Signature Scheme

The 3WISE digital signature scheme is specified in [6]. The construction is a small field multivariate cryptosystem with the normal construction given in Figure 1. There is a central nonlinear polynomial map F which is perturbed by two linear transformations: one acting on the inputs and one acting on the outputs. The central map must be specially structured so that it is easily inverted. The hope is that the public map P is hard to invert.

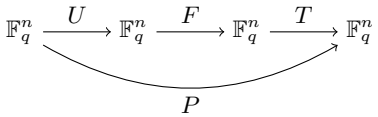

Fig. 1. Generic construction of a small field multivariate cryptosystem.

In the case of 3WISE, the central map F is extremely simple. It is coordinate-wise the monomial map $x \mapsto x^3$. Naturally the field $\mathbb{F}_q$ is chosen so that this map is non-linear and invertible. The 3WISE specifications require $q = 17$, but any field $\mathbb{F}_q$ for which $q-1$ and 3 are coprime (and $x \neq x^3$) can be chosen.

To sign the hash of a message encoded into $\mathbb{F}_q^n$, say $\mathbf{y}$, requires an application of the inverse of each of the three component maps. Inverting F is simple because F^{-1} is also coordinate-wise a monomial map. Specifically each coordinate is raised to the power of $3^{-1} \pmod{q-1}$. Thus the signature is given by the vector $\mathbf{x} = U^{-1}(F^{-1}(T^{-1}(\mathbf{y})))$

Verification is accomplished by computing $\mathbf{y}$, the encoding of the hash of the message, and evaluating P at the signature $\mathbf{x}$. Verification succeeds with probability one because

$$P(\mathbf{x}) = T(F(U(U^{-1}(F^{-1}(T^{-1}(\mathbf{y})))))) = \mathbf{y}.$$

3 MinRank

The MinRank problem is the problem of determining a linear combination of a given collection of matrices that satisfies a rank bound. Specifically, given K matrices $\mathbf{M}_i$ of dimension $m \times n$ with coefficients lying in a field $\mathbb{F}$, and a target rank r, the MinRank problem asks us to recover a nonzero collection of coefficients λ_i lying in the field $\mathbb{E}$ such that $\sum_i \lambda_i \mathbf{M}_i$ has rank bounded by r. In general the field $\mathbb{E}$ may be a subfield or an extension of $\mathbb{F}$.

There are numerous ways of computing a linear combination satisfying the requirements of a MinRank solution. Such methods include brute force, linear algebra search [7], Kipnis-Shamir modeling [10], minors modeling [5], and support minors modeling [3].

For our application in this manuscript our target rank is 1 and we are considering square matrices, so the Kipnis-Shamir and support minors models are not competitive, having the same solving degree as minors modeling but with many more variables. The brute force method is only viable when the number of matrices K and the size of the field $\mathbb{E}$ are small, which is also not relevant here. Thus, we will focus on the linear algebra search and minors modeling approaches. Since we are interested in MinRank instances with n matrices of dimension $n \times n$, we restrict to this case in the descriptions below.

3.1 Linear Algebra Search

Since matrices of lower rank have larger kernels, a randomly chosen vector has a greater probability of being in the kernel of a lower rank matrix in comparison to a higher rank matrix. The linear algebra search method, see [7], takes advantage of this fact to provide a combinatorial search problem that is usually more efficient than brute force.

Specifically, one chooses a random vector $\mathbf{x}$ and constructs the system of linear equations

$$\sum_{i=1}^{n} t_i \mathbf{x} \mathbf{M}_i = \mathbf{0}$$

in the unknowns t_i. This system is fully determined, and so there is a small solution space in general. One then constructs the matrix

$$\mathbf{L} = \sum_{i=1}^{n} t_i \mathbf{M}_i,$$

and checks to see if the rank condition is met.

Given a target rank of r, the probability that a randomly selected vector is in the kernel is q^{-r}, so one expects for the process to need to be repeated q^r times. The total complexity of the method is then

$$\mathcal{O}\left(q^r n^\omega\right),$$

where ω is the linear algebra constant.

3.2 Minors Modeling

Another technique for solving MinRank is the minors method of [5]. If there is a linear combination of the matrices $\mathbf{M}_1, \ldots, \mathbf{M}_n$ of rank at most r, then all $(r+1) \times (r+1)$ minors of the sum

$$\mathbf{\Sigma} = \sum_{i=1}^{n} x_i \mathbf{M}_i$$

are zero for the correct values of the x_i. Thus, we may take these minors as a system of equations and solve it.

For most MinRank instances occuring in multivariate cryptography, the parameters of the minors modeling instance are such that the system of degree $r+1$ equations linearizes and can be solved immediately by specifying a variable. In this case, the complexity of solving such a MinRank instance with minors modeling is

$$\mathcal{O}\left(\binom{n+r}{r+1}^\omega\right).$$

4 Attacks on 3WISE

In this section we present a couple of attacks that break the proposed parameters for 3WISE in [6]. The first attack is specific to the parameter set used and can be made inefficient by changing the parameters. The second attack, based on rank, is more intrinsic to the design of the scheme and breaks all conceivable parameters.

Before we present either of these attacks, we should note that the parameters do not achieve the claimed NIST security levels due to targeting the wrong values by the designer of the scheme. NIST level 1, for example, is defined as providing as much security as brute force key search for AES-128, under an assumption that AES-128 works as an ideal cipher. As specified in [8,9], it is assumed that one call of AES-128 on a guessed key costs around 2^{15} gates. Thus NIST level I requires 2^{143} gates (not 2^{128} gates) in the absence of some other justified cost, such as memory access. Since an essentially memoryless brute force attack for the parameters in [6] require fewer gates, they do not achieve their claimed security levels as defined.

4.1 First Attack: Interpolation

The first observation providing a substantial break of 3WISE is based on the fact that the inverse of the public key exists and is a low degree polynomial map. Specifically, with the selection of $q = 17$, the inverse of the map $f : \mathbb{F}_q \to \mathbb{F}_q$ defined by $f(x) = x^3$ is $f(x) = x^{11}$. Since U and T are linear, it is clear that $P^{-1} = U^{-1} \circ F^{-1} \circ T^{-1}$ is a degree 11 map.

As in [2], we can interpolate the inverse of the map by simply generating a sufficiently large number of plaintext/ciphertext pairs. Since a generic degree 11 polynomial in n variables has $\binom{n+10}{11}$ monomials, we can find the inverse of the public key as a polynomial with complexity

$$\mathcal{O}\left(\binom{n+10}{11}^{\omega}\right) = \mathcal{O}\left(n^{11\omega}\right)$$

as measured in field multiplications. Once recovered, this inverse map is a new (though very large) decryption key.

The complexity of this calculation is summarized in Table 1. Although the parameters were mistakenly chosen, they were roughly correct in terms of the brute force complexity. With respect to the interpolation attack, however, none of the parameters approach their claimed security levels.

Table 1. Complexity (logarithmic) of the Inverse Interpolation Attack on 3WISE for all parameter sets.

3WISE(q, n)	Claimed Security	Brute Force	Interpolation
$2WISE(17, 32)$	2 (2^{146} gates)	131	95
$3WISE(17, 48)$	4 (2^{210} gates)	196	111
$3WISE(17, 64)$	5 (2^{272} gates)	262	123

The damage to the scheme due to the interpolation attack is easily mitigated by the same method that fixes the problem due to the brute force attack. The parameter q should be increased. If the user selects $q = 257$ instead of $q = 17$, then the brute force attack becomes quite bad, and since the inverse of 2 modulo

256 is 171, the inverse polynomial has $\binom{n+170}{171}$ monomials, and the attack is completely impractical.

Naturally, since the recovered inverse is quite large in terms of storage size, one could argue that actually computing with the inverse is very costly in terms of memory access. Even using a very conservative memory access model in which memory accesses cost the square root of the memory size (and under the unrealistic assumption that recovering a single coefficient of the polynomial is a random access) the cost of evaluating the inverse polynomial is well below 2^{100} gate-equivalents for every parameter set.

4.2 Second Attack: MinRank

The second and most devestating attack on 3WISE is an attack powered by MinRank. This attack breaks all conceivable parameter sets significantly.

First, note that since the central polynomials are all of the form $F_i(\mathbf{x}) = x_i^3$, they can all be modeled as 3-tensors of rank 1. Specifically the polynomial F_i can be represented as the 3-tensor $\mathbf{F}_i$ whose value is $1 \in \mathbb{F}_q$ on the basis vector $e_i \otimes e_i \otimes e_i$ and is zero on all other basis vectors. (Visually, this 3-tensor is a cube of zero coefficients with a single 1 at coordinate (i, i, i).)

Clearly, the composition $\mathbf{F}_i \circ U$ corresponds to the 3-tensor $\mathbf{F}_i(U\cdot, U\cdot, U\cdot)$. The public polynomials then are linear combinations of these 3-tensors.

The key observation is the fact that the property that the private 3-tensors have rank 1 implies that any projection onto a 2-tensor must also have rank 1. Specifically, for any $\widetilde{\mathbf{z}} \in \mathbb{F}_q^n$, we have that $\mathbf{F_i}(\widetilde{\mathbf{z}}, \cdot, \cdot)$ is a 2-tensor of rank 1. In fact, we can write $\mathbf{F}_i(\widetilde{\mathbf{z}}, \cdot, \cdot)$ as a matrix with a single potentially nonzero entry, $\widetilde{z}_i$ at coordinate (i, i).

Composing by U does not change the rank of this matrix. We simply exchange $\widetilde{\mathbf{z}}$ with $U(\mathbf{z})$, and place U in the other coordinates. Thus $\mathbf{F}_i(U\mathbf{z}, U\cdot, U\cdot)$ is a rank 1 matrix for any choice of $\mathbf{z}$. Since the public cubic forms are linear combinations of the 3-tensors $\mathbf{F}_i(U\cdot, U\cdot, U\cdot)$ and the operations of projection and taking linear combinations of these 3-tensors commute, we see that there are n linearly independent linear combinations of the public 3-tensors with the property that any projection by a vector $\mathbf{z}$ produces a rank 1 matrix.

If there were an exact correspondence between rank 1 matrices in the span of $\mathbf{P}_i(\mathbf{z}, \cdot, \cdot)$ and the 3-tensors $\mathbf{F}_i(U\cdot, U\cdot, U\cdot)$ then it would be enough to just solve the MinRank problem. Unfortunately, the MinRank solution is not enough information to recover the rank 1 matrices $\mathbf{F}_i(U\mathbf{z}, U\cdot, U\cdot)$. To see this fact, note that

$$\sum_{i=1}^{n} t_i \mathbf{P}_i(\mathbf{z}, \cdot, \cdot) = \sum_{j=1}^{n} z_j \left(\sum_{i=1}^{n} t_i \mathbf{P}_i(\mathbf{e}_j, \cdot, \cdot) \right). \tag{1}$$

If the sum $\sum_{i=1}^{n} t_i \mathbf{P}_i(\cdot, \cdot, \cdot)$ has low rank as a 3-tensor, it is more likely that there is a choice of $\mathbf{z}$ making the right hand side of Equation (1) of rank 1. Since there are very many low rank 3-tensors in the span of the $\mathbf{F}_i$ and hence the $\mathbf{P}_i$, the probability that the solution $\mathbf{t}$ to the MinRank instance produces $\sum_{i=1}^{n} t_i \mathbf{P}_i(\mathbf{z}, \cdot, \cdot)$ of low (but not 1) rank is fairly high.

To combat this problem we may select multiple projections $\mathbf{z}_j$ and require that simultaneously each of the $\sum_{i=1}^{n} t_i \mathbf{P}_i(\mathbf{z}_j, \cdot, \cdot)$ are of rank 1. This requirement is equivalent to having the multiple conditions

$$\text{Rank}\left(\sum_{k=1}^{n} z_{j,k}\left(\sum_{i=1}^{n} t_i \mathbf{P}_i(\mathbf{e}_k, \cdot, \cdot)\right)\right) = 1$$

on the 3-tensor $\sum_{i=1}^{n} t_i \mathbf{P}_i(\cdot, \cdot, \cdot)$. Experimentally, we need only solve this simultaneous MinRank problem with three random vectors $\mathbf{z}_j$ to eliminate the spurious solutions.

This "simultaneous MinRank problem" is actually just a special case of normal MinRank. Instead of the square matrices $\mathbf{P}_i(\mathbf{z}_j, \cdot, \cdot)$, for $i = 1, \ldots, n$, we consider the concatenations $\mathbf{P}_i(\mathbf{z}_1, \cdot, \cdot) \| \mathbf{P}_i(\mathbf{z}_2, \cdot, \cdot) \| \mathbf{P}_i(\mathbf{z}_3, \cdot, \cdot)$ for $i = 1, \ldots, n$.

To recover the inverse of the output transformation for an equivalent key, we must recover a linearly independent collection of n MinRank solutions. The solutions then form the rows of the matrix representation of a linear transformation $\widetilde{T}^{-1}$.

Once the map $\widetilde{T}^{-1}$ is recovered, the scheme unravels quickly. Notice that if $\mathbf{y} = P(\mathbf{x})$, then $\mathbf{v} = \widetilde{T}^{-1}(\mathbf{y})$ is some permutation of nonzero multiples of the coordinates of $F(U(\mathbf{x}))$. Thus, taking the 11th power of each coordinate v_i produces a nonzero multiple of some coordinate u_j of $U(\mathbf{x})$, since $F_i^{-1}(x) = x^{11}$. Therefore $F^{-1}(\widetilde{T}^{-1}(\mathbf{y}))$ is linearly related to $\mathbf{x}$. To recover an equivalent input transformation $\widetilde{U}$, we merely generate sufficiently many certificate/signature pairs and solve. We then obtain an equivalent key

$$\widetilde{T} \circ F \circ \widetilde{U} = T \circ F \circ U.$$

5 Experiments and Complexity Analysis

We studied the two possibly best MinRank methods for this attack to determine their relative performance: linear algebra search and minors modeling. We note that with minors modeling the system always linearizes at degree 2, so the attack has good performance. The greatest factor in complexity for both of these methods is the fact that we need to perform MinRank on the order of n times.

The complexity of the linear algebra search method on a single MinRank instance is $\mathcal{O}(q^r n^{\omega+1})$ due to the need for n matrix multiplications to construct the coefficient matrix and solve for $\mathbf{t}$. With the simultaneous instance we are using in the attack, the situation is a bit complicated. We require many more conditions on the unknowns than for a square instance, but because of their related nature, the systems are often consistent and provide solutions.

The analysis of the minors method is much simpler. Since the system linearizes at degree 2, we only need to compute roughly $\binom{n+1}{2}$ minors and then solve the linear system to recover the coefficients. The complexity of this task is $\mathcal{O}(\binom{n+1}{2}^{\omega})$. Performing this task n times produces a total complexity of the dominant step of $\mathcal{O}(n\binom{n+1}{2}^{\omega})$.

We implemented both MinRank methods to attack the scheme in the MAGMA Computer Algebra System[3], see [4]. We performed these experiments on the all proposed parameters from the specification [6] .

In all instances, for both methods, the attack worked in practice to break the scheme. The results are summarized in Table 2.

Table 2. MAGMA attack timing for 100 instances of the 3WISE digital signature scheme for parameters claiming NIST level 2, 4 and 5 parameters.

		Lin. Alg. Search		
3WISE(q, n)	Sec. Level	Least(ms)	Average(ms)	Most(ms)
3WISE$(17, 32)$	2	1960	2315	2850
3WISE$(17, 48)$	4	13900	17534	47120
3WISE$(17, 64)$	5	38120	59104	83040
	Minors			
3WISE(q, n)	Sec. Level	Least(ms)	Average(ms)	Most(ms)
3WISE$(17, 32)$	2	720	1219	1620
3WISE$(17, 48)$	4	5460	7017	15990
3WISE$(17, 64)$	5	18550	20392	25390

We should note explicitly that the first step of the attack, the MinRank step, depends only poly-logarithmically on q, the field size. So the first step of the attack works the same way with minors modeling and with essentially the same complexity with any value of q. Furthermore, the attack works for even higher exponents than 3 by choosing projections down to 2-tensors. Thus, there is no hope in protecting the scheme even if one were willing to have completely unrealistic parameters.

6 Conclusion

The 3WISE scheme is an original attempt to produce a secure multivariate digital signature scheme. One could consider it like a version of the famous C^* scheme, see [11], but with the power map working on a different $\mathbb{F}_q$ algebra than an extension field and with a cubic key instead of quadratic.

Interestingly, even though the key is cubic, we can apply standard MinRank techniques to break the scheme completely. We verify that not only are the proposed parameters completely and practically broken, but that there is no set of parameters that will secure 3WISE in its current state without some sort of modification.

[3] Certain equipment, instruments, software, or materials, commercial or non-commercial, are identified in this paper in order to specify the experimental procedure adequately. Such identification is not intended to imply recommendation or endorsement of any product or service by NIST, nor is it intended to imply that the materials or equipment identified are necessarily the best available for the purpose.

References

1. Alagic, G., Apon, D., Cooper, D., Dang, Q., Dang, T., Kelsey, J., Lichtinger, J., Liu, Y.K., Miller, C., Moody, D., Peralta, R., Perlner, R., Robinson, A., Smith-Tone, D.: Status report on the third round of the NIST post-quantum cryptography standardization process. Tech. Rep. NIST Interagency or Internal Report (IR) 8413, National Institute of Standards and Technology, Gaithersburg, MD (July 2022). https://doi.org/10.6028/NIST.IR.8413-upd1
2. Alperin-Sheriff, J., Ding, J., Petzoldt, A., Smith-Tone, D.: Total break of the fully homomorphic multivariate encryption scheme of 2017/458: Decryption can not be of low degree. IACR Cryptol. ePrint Arch. p. 471 (2017), http://eprint.iacr.org/2017/471
3. Bardet, M., Bros, M., Cabarcas, D., Gaborit, P., Perlner, R.A., Smith-Tone, D., Tillich, J., Verbel, J.A.: Improvements of algebraic attacks for solving the rank decoding and minrank problems. In: Moriai, S., Wang, H. (eds.) Advances in Cryptology - ASIACRYPT 2020 - 26th International Conference on the Theory and Application of Cryptology and Information Security, Daejeon, South Korea, December 7-11, 2020, Proceedings, Part I. Lecture Notes in Computer Science, vol. 12491, pp. 507–536. Springer (2020). https://doi.org/10.1007/978-3-030-64837-4_17, https://doi.org/10.1007/978-3-030-64837-4_17
4. Bosma, W., Cannon, J., Playoust, C.: The magma algebra system i: The user language. J. Symb. Comput. **24**(3–4), 235–265 (Oct 1997). https://doi.org/10.1006/jsco.1996.0125, https://doi.org/10.1006/jsco.1996.0125
5. Faugère, J., Levy-dit-Vehel, F., Perret, L.: Cryptanalysis of minrank. In: Wagner, D.A. (ed.) Advances in Cryptology - CRYPTO 2008, 28th Annual International Cryptology Conference, Santa Barbara, CA, USA, August 17-21, 2008. Proceedings. Lecture Notes in Computer Science, vol. 5157, pp. 280–296. Springer (2008). https://doi.org/10.1007/978-3-540-85174-5_16, https://doi.org/10.1007/978-3-540-85174-5_16
6. Gómez Rodríguez, B.: 3wise: Cubic element-wise trapdoor based mpkc cryptosystem. NIST CSRC (2023), https://csrc.nist.gov/Projects/pqc-dig-sig/round-1-additional-signatures
7. Goubin, L., Courtois, N.T.: Cryptanalysis of the TTM cryptosystem. In: Okamoto, T. (ed.) Advances in Cryptology - ASIACRYPT 2000, 6th International Conference on the Theory and Application of Cryptology and Information Security, Kyoto, Japan, December 3-7, 2000, Proceedings. Lecture Notes in Computer Science, vol. 1976, pp. 44–57. Springer (2000). https://doi.org/10.1007/3-540-44448-3_4, https://doi.org/10.1007/3-540-44448-3_4
8. Group, C.T.: Submission requirements and evaluation criteria for the post-quantum cryptography standardization process. NIST CSRC (2016), http://csrc.nist.gov/groups/ST/post-quantum-crypto/documents/call-for-proposals-final-dec-2016.pdf
9. Group, C.T.: Call for additional digital signature schemes for the post-quantum cryptography standardization process. NIST CSRC (2022), https://csrc.nist.gov/csrc/media/Projects/pqc-dig-sig/documents/call-for-proposals-dig-sig-sept-2022.pdf
10. Kipnis, A., Shamir, A.: Cryptanalysis of the hfe public key cryptosystem by relinearization. Advances in Cryptology - CRYPTO 1999, Springer **1666**, 788 (1999)
11. Matsumoto, T., Imai, H.: Public quadratic polynominal-tuples for efficient signature-verification and message-encryption. In: Günther, C.G. (ed.) Advances

in Cryptology - EUROCRYPT '88, Workshop on the Theory and Application of of Cryptographic Techniques, Davos, Switzerland, May 25-27, 1988, Proceedings. Lecture Notes in Computer Science, vol. 330, pp. 419–453. Springer (1988). https://doi.org/10.1007/3-540-45961-8_39, https://doi.org/10.1007/3-540-45961-8_39